Let's meet as usual: Do games played on non-frequent days differ? Evidence from top European soccer leagues

Daniel Goller and Alex Krumer*

Swiss Institute for Empirical Economic Research
University of St.Gallen
Molde University College
Specialized University in Logistics

First version: May 2019; This version: November 2019

Abstract: Balancing the allocation of games in sports competitions is an important organizational task that can have serious financial consequences. In this paper, we examine data from 10,142 soccer games played in the top German, Spanish, French, and English soccer leagues between 2007/2008 and 2016/2017. Using a machine learning technique for variable selection and applying a semi-parametric analysis of radius matching on the propensity score, we find that all four leagues have a lower attendance in games that take place on non-frequently played days than those on frequently played days. We also find that, in all leagues, there is a significantly lower home advantage for the underdog teams on non-frequent days. Our findings suggest that the current schedule favors underdog teams with fewer home games on non-frequent days. Therefore, to increase the fairness of the competitions, it is necessary to adjust the allocation of the home games on non-frequent days in a way that eliminates any advantage driven by the schedule. These findings have implications for the stakeholders of the leagues, as well as for coaches and players.

Keywords: (D) OR in sports; Performance; Schedule effects; Soccer.

JEL classification: D00, L00, D20, Z20.

Address for correspondence:
Alex Krumer, Faculty of Business Administration and Social Sciences, Molde University College, Britvegen 2, Molde, 6402, Norway, alex.krumer@himolde.no.
Daniel Goller (corresponding author), Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstrasse 14, CH-9000 St. Gallen, Switzerland, daniel.goller@unisg.ch.

* We would like to thank Michael Lechner, Michael Knaus, Gabriel Okasa and Sandro Heiniger for helpful comments and suggestions. The usual disclaimer applies.

1 Introduction

In recent decades, top European soccer leagues have become large business corporations. Each of the top leagues receives more than 1 billion Euros from television revenues alone.[1] A large part of these amounts is redistributed to teams based on their performance. In addition, the highest ranked teams of the top leagues earn the right to participate in the UEFA Champions League and Europa League. According to UEFA (the governing body of soccer in Europe), in the 2016/2017 season, more than 1.3 billion Euros was shared among the clubs in the Champions League and almost 400 million Euros among the clubs in the Europa League.[2] Since an unbalanced schedule may have serious financial consequences, the leagues face an important organizational task in creating a schedule that will not discriminate against or favor specific teams.

Top European soccer leagues use a double round-robin structure, where each team competes against each other team twice during the season. Operational research literature has intensively investigated different issues of round-robin structures, such as balanced distribution of home and away matches (Della Croce and Oliveri, 2006; Durán, Guajardo and Sauré, 2017), break optimization (Ribeiro and Urrutia, 2007), police requirements (Kendall et al., 2010), stakeholders' requirements (Goossens and Spieksma, 2009), minimizing traveling distance (Kendall, 2008; Durán et al., 2019), optimizing the number of prizes (Krumer, Megidish and Sela, 2019), and allocation of rescheduled games (Yi, Goossens and Nobibon, 2019).[3] However, the operational research literature has neglected another important issue: the

[1] From https://www.euronews.com/2018/05/30/footbal-broadcast-rights-ligue-1-championship-booms-into-the-billion and https://www.soccerex.com/insight/articles/2018/laliga-s-new-tv-rights-distribution-model-a-level-playing-field. Last accessed on 31.01.2019.

[2] From http://www.uefa.com/uefachampionsleague/news/newsid=2398575.html and http://www.uefa.com/uefaeuropaleague/news/newsid=2398584.html . Last accessed on 31.01.2019.

[3] See the comprehensive reviews of Rasmussen and Trick (2008), Kendall et al. (2010), Goossens and Spieksma (2012), and Van Bulck et al. (2020) for additional references.

allocation of (non-rescheduled) games between days that are not the usual days in a league's calendar. This may play an important role, because fans may have different preferences toward certain days of the week (Wang, Goossens and Vandebroek, 2018) or even the timings of the game (Krumer, 2019). For example, if fans are not used to attending games on a certain day, their routine may prevent them from attending games on those days. In such cases, lower attendance may be expected on these days, which may reduce the home advantage (Downward and Jones, 2007; Nevill et al., 2002; Page and Page, 2010; Pettersson-Lidbom and Priks, 2010). In the 2017/2018 season, for example, the decision to schedule Monday games in the German Bundesliga 1 led to large protests. Moreover, German team 1.FSV Mainz 05 officially complained to the German Federation (DFL) since it had to play eight non-weekend games, six of which were at home. The DFL finally decided to abolish Monday games.[4]

The present paper is closely related to the study of Krumer and Lechner (2018), who investigated games in the German Bundesliga 1 and found a significantly lower attendance and also a lower home advantage in midweek days compared to weekend days (Friday, Saturday, Sunday − the most frequently played days in this league). In other leagues, however, the three most frequent days, which account for approximately 90 percent of all matches, differ from those in Bundesliga 1. For example, in England and France, the three most frequent days for games are Saturday, Sunday, and Wednesday, whereas the respective days in Spain are Saturday, Sunday, and Monday.[5]

[4] For additional information, see: https://www.mainz05.de/news/brief-an-die-dfl-kritik-an-terminierungen/ (in German) and https://www.dw.com/en/bundesliga-monday-games-to-be-discontinued-as-fan-protests-persist/a-46390559. Last accessed on 03.11.2019.

[5] According to the UEFA association club coefficients, these are four of the five most successful leagues in Europe. We do not use data on the fifth (the Italian Serie A) since it suffered from various scandals and club insolvencies in the underlying period. See, for example, Buraimo, Migali and Simmons (2016), who found a significantly lower crowd attendance after the Calcipoli scandal in the 2005/2006.

In this paper, we ask a simple question: Does playing on non-frequent days have any effect on the various aspects of soccer games? More specifically, using data from the four above-mentioned European soccer leagues between 2007/2008 and 2016/2017, we compare the games that were played on frequent and on non-frequent days with regard to their attendance and home advantage. More specifically, unlike Krumer and Lechner (2018), we separately investigated games with a home advantage for the favorite team and games with a home advantage to the underdog team. This is another contribution to the operational research literature, which has not yet considered the possibility of such a heterogeneous effect.[6] To the best of our knowledge, the only paper to study such an effect in a context of schedule is Krumer (2019), which investigated the UEFA Europa League games with kick-off times at 19:00 CET and 21:05 CET. That paper documented a lower attendance in games that started at 21:05 CET and a significantly lower home advantage for the underdog teams in these later games.

It is important to note that the allocation of the match days is not entirely random, and might be based on different schedule-related features such as public holidays, international breaks, European tournaments, police requirements, broadcasters' and clubs' interests, months of the year, and even teams' values. Therefore, we need to control for these deviations from random selection into treatment (that is, non-frequent days) using a selection-on-observables approach. Specifically, we estimated the average treatment effect of playing on the non-frequent days by using the distance-weighted radius matching approach with bias adjustment suggested by Lechner, Miquel, and Wunsch (2011). This estimator is constructed to be more robust than other matching-type estimators, as it combines the features of distance-weighted radius matching with a bias adjustment to remove sample biases due to mismatches (Huber, Lechner and Wunsch, 2013). In addition, having a rich database in terms of potential confounding

[6] We discuss this type of heterogeneity in detail in the data section.

variables, we use a machine learning technique for variable selection as proposed by Belloni, Chernozhukov, and Hansen (2014).

Based on analysis of 10,142 games from the top four European leagues over 10 seasons, we found a significantly lower attendance on non-frequent days in all four leagues. In addition, all of the leagues had a reduced home advantage on non-frequent days for the underdog teams, which is in line with Krumer (2019). Our results suggest that the difference in the number of points between the favorite and the underdog teams, when the game takes place on non-frequent days compared to frequent days, is 0.49 in Ligue 1, 0.43 in La Liga, 0.42 in the Premier League, and 0.63 in Bundesliga 1. To put these numbers into perspective, a favorite with home advantage gains about 1.1 points more than the underdog, on average.[7]

Such a reduced home advantage for weaker teams in games with a lower attendance is in line with the literature on the effect of the density of the crowd and its noise on referees' bias in favor of the home team. For example, using laboratory settings, Nevill et al. (2002) determined that crowd noise had a significant effect on the probability of a referee issuing a yellow card against a home team. Downward and Jones (2007) showed a positive relationship between the size of the crowd and the likelihood of a player receiving a yellow card in the English FA Cup. Similarly, Pettersson-Lidbom and Priks (2010) found a significant home bias of referees in games in which spectators were present compared to games with no spectators at all in the Italian Serie A. In addition, Ponzo and Scoppa (2018) showed a significantly larger number of cards against away teams in Serie A games between the teams from the same city that shared the same stadium. Finally, Page and Page (2010) found that the home advantage effect differs significantly among referees, and that this relationship is moderated by the size of the crowd. Therefore, a possible mediator of the difference in the home advantage of the

[7] Note that the winning team receives three points, while the losing team gets no points. In case of a draw, each team gets one point.

underdog teams in games that take place on non-frequent days is lower crowd noise compared to games on frequent days. However, except for the case of Bundesliga 1, whose favorite teams that played at home suffer from the highest reduction of the crowd on non-frequent days (almost 16 percent smaller crowds), we found no difference in home advantage between different days when favorite teams play at home. As Krumer (2019) proposed, it is possible that the underdog teams depends more on crowd support than the favorite teams because the latter are likely to win due to their higher ability regardless of home support. Therefore, underdog teams seem to lose more points in games with lower crowd density compared to the favorite teams.

Our results suggest that since some underdog teams play more home games on non-frequent days than other underdog teams, the current structure favors underdogs that play fewer home games on non-frequent days and favorites that play more away games on non-frequent days. To illustrate a possible relationship between an unbalanced schedule and the resulting monetary rewards, Krumer and Lechner (2018) gave an example of *SC Paderborn 07*, which was relegated from the Bundesliga 1 in the 2014/2015 season. This team played more home games on non-frequent days than its closest rival until the very last game in the relegation fight, *Hamburger SV*, which eventually remained in the top division. Moreover, one of these games was against *Hamburger SV*, which the latter won. According to Krumer and Lechner (2018), if *SC Paderborn 07* had survived in the Bundesliga 1, its additional revenue from TV alone would have been at least 10.3 million Euros (not counting all other revenues from ticketing, advertising, and so on).

The remainder of the paper is organized as follows. Section 2 describes the schedule of the different leagues. The data and some descriptive results are presented in Section 3. Section 4 presents the empirical strategy. The results are contained in Section 5 and we offer concluding remarks in Section 6.

2 Description of the leagues' schedules

2.1 General structure of the leagues

While there are specific features for different leagues, the structure of all four leagues we investigate is largely similar. The leagues are organized as double round-robin tournaments, with each round consisting of $\frac{n}{2}$ games, where n is the number of teams in the league. In total, each team plays each other team twice, once at its home field in the first half of the season, and once away in the second half of the season (or vice versa). In total, every team has $2(n-1)$ games. In the French, Spanish, and English leagues, there are 20 teams, resulting in 38 games for each team. In the German Bundesliga 1, there are 18 teams, resulting in 34 games for each team. In addition, except for the English Premier League, the leagues have a winter break of several weeks without games.

The schedule of the leagues should also take into account international tournaments between nations, with the requirement to release participating players earlier and allow them a longer vacation. The main tournaments are the FIFA World Cup and the UEFA European Championship (held alternately every two years in June and July). Other tournaments that have the requirement to release players are the African Cup of Nations and the Asian Cup. Those take place during wintertime in parallel to the European leagues' matches.

League games usually take place on weekends, but since there are not enough weekends in the season, some rounds take place on other days.

At the end of a season, the final table determines which teams participate in the following season's European club tournaments; these include the Champions League, which is the most prestigious club tournament in Europe, and the Europa League, which also yields significant monetary rewards. In addition, the two or three worst-ranked clubs are relegated to the second

division, implying that the different outcomes have substantial financial consequences for the clubs.

Following Krumer and Lechner (2018), who investigated the effect of playing on Fridays, Saturdays, and Sundays (the three most frequently played days in the Bundesliga 1), we identified the three most frequently played days separately for each league, as depicted in Table 1. Although less pronounced than in the Bundesliga 1, the three most frequent days for Ligue 1 and Premier League are quite clear. For the La Liga, however, the reduction in frequency between the third and fourth days is less significant. Nevertheless, we have stuck to the choice of Krumer and Lechner (2018) by taking the three most frequent days and need the remaining four days, which are defined as non-frequent, to have enough observations. In addition, we discuss special settings and uniqueness of schedule of the games that are described below for each league separately.

[Insert Table 1 here]

2.2 The French Ligue 1

The three most frequently played match days are Saturday, Sunday, and Wednesday. The seasonal tournament in France takes place from August to the beginning of May. The top three teams advance to the Champions League (or for the Champions League playoffs). Teams in the fourth to sixth positions play in the Europa League (this may also depend on the outcome of an elimination French Cup tournament, called the *Coupe de France*). In addition, the two worst-ranked clubs are relegated to the lower division and the 18th-ranked team has to participate in a relegation playoff against the team that won the second division playoff for the right to play in the Ligue 1 in the following year.[8]

[8] Prior to 2016/17, the three bottom-ranked teams were directly relegated to the lower division.

2.3 The Spanish La Liga

The three most frequently played match days are Saturday, Sunday, and Monday. The seasonal tournament in Spain runs from the end of August or beginning of September until May of the following year. The top four teams advance to the Champions League (or the playoffs). Teams finishing fifth to seventh play in the Europa League (this may also depend on the outcome of an elimination Spanish Cup tournament, called the *Copa del Rey*). In addition, the three worst-ranked clubs are relegated to the lower division.

2.4 The German Bundesliga 1

The three most frequently played match days are Saturday, Sunday, and Friday. The seasonal tournament in Germany takes place from August to May. The top four teams advance to the Champions League (or to the playoffs). Teams finishing fifth to seventh play in the Europa League (this may also depend on the outcome of an elimination German Cup tournament, called the *DFB-Pokal*). In addition, the two worst-ranked clubs are relegated to the lower division and the 16th-ranked team must participate in the relegation playoffs against the third-ranked team in the Bundesliga 2 for the right to play in the Bundesliga 1 in the following year.[9]

2.5 The English Premier League

The three most frequently played match days are Saturday, Sunday, and Wednesday. The seasonal tournament in England takes place from August until May. This is the only one of the four leagues discussed here that does not have a long winter break. Several rounds take place during the Christmas holidays, usually involving local derbies to avoid fans having to travel long distances on those days (Kendall, 2008). The best-known round takes place on Boxing

[9] The relegation playoff format was introduced in the 2008/2009 season. Prior to that, three teams were directly relegated to the Bundesliga 2.

Day, which is a part of the Commonwealth tradition. We expect this to play a role in the scheduling process in the underlying period and account for this issue, as described in the next section.

The top four teams advance to the Champions League (or the playoffs). Teams finishing fifth to seventh play in the Europa League (this may also depend on the outcome of two elimination English Cup tournaments: the FA Cup and the League Cup). In addition, the three worst-ranked clubs are relegated to the lower division.

Compared to the other three leagues, the Premier League has the highest amount of rescheduled games, because its clubs potentially have the highest number of games to play in their national cups (the FA Cup and the League Cup). The reason for this is that, in most stages of these competitions, a drawn match necessitated a repeated second game.[10] This partly interfered with the initial schedule proposed by the calendar committee.

3 Data and descriptive results

3.1 Database

We used data on four major European football leagues: the French Ligue 1, German Bundesliga 1, the Spanish La Liga, and the English Premier League.[11] For each of the leagues, we collected data on all the games starting from the start of the 2007/2008 season until the end of the 2016/2017 season. This represents a total of 14,460 games. However, we disregarded games in which a home team did not play at its usual home stadium. For example, Bayer Leverkusen from Germany did not play the second half of the 2008/2009 season at its home stadium due to reconstruction. RC Lens from France experienced a similar situation in the 2014/2015 season. In addition, Montpellier, Caen (both 2014/2015), and Lille (2007/2008 and

[10] In the case of a draw after the second game, overtime is played and, if needed, penalty shootouts determine the winner.

[11] See Appendix C for the full list of sources.

2008/2009) from France played some home games in alternative stadiums. We also removed matches in which one of the teams had already been relegated or had already won the championship title.[12] In addition, teams that play in the Champions League or Europa League may strategically adjust their squads in the domestic leagues games that take place just before or after the European cups (for example, they may save their best players before the European games to avoid a risk of injury or let them rest after). Therefore, we also removed games that involved teams playing just before or just after the continental competitions.[13] Finally, we also removed rescheduled games, since those may differ with regard to media attention as they are detached from the rest of the matches.[14] Removing those games left 10,142 matches, 9,010 of which took place on frequent days and 1,132 on non-frequent days.

For every game we collected information on the identity of teams, referees, exact day, attendance, distance between the cities, and the final score. We also used data from the *Transfermarkt* website to proxy the market value of each player of each team in every season. This data also includes personal information of each player, such as his age, height, and preferred foot. Finally, we have data on the dates of the beginning and the end of each coach's tenure, as well as data on the capacity of each stadium.

3.2 Definition of heterogeneity

There can be different types of heterogeneity in sports competition, such as home versus away or the favorite versus underdog teams.[15] We choose the favorite-underdog type of

[12] For example, in 2013/2014 season Bayern Munich from Germany had won the Bundesliga 1 title after 27 rounds. However, in the next three games they only gained one point out of nine and were accused of lacking motivation. See Kendall and Lenten (2017) for additional discussion on the usage of squads in remaining games after winning a title.

[13] See Rohde and Breuer (2017), who showed that teams adjust their efforts in domestic league just before or after games in European tournaments.

[14] For additional discussion on rescheduled games, see Yi, Goossens, and Nobibon (2019).

[15] Other types of heterogeneities might be found in teams that replaced their coach, matches played on artificial versus natural grass, televised versus non-televised matches, traditional rivalry versus non-rivalry matches, etc.

heterogeneity because it is intuitive that probabilities of winning (or the expected number of points) are largely driven by the differences in the teams' abilities, whereas the home-away factor plays a secondary role in increasing or decreasing the gap between the teams' probabilities of winning. While home advantage is a well-established phenomenon, the literature has largely neglected the heterogeneous effect for favorites versus underdogs. More importantly, beyond the above-mentioned intuition, standard economic theory predicts probabilities of winning based on contestants' innate abilities. For example, the Tullock contest (Tullock, 1980) is a well-known model in economic theory that has been applied in many fields, from political races (e.g. Klumpp and Polborn, 2006) to sports tournaments (e.g., Szymanski, 2003). The most popular versions of this model are lottery and all-pay contest. In the lottery version, a contestant with a lower effort still has a positive probability of winning, whereas an all-pay contest is fully discriminatory, where a contestant with a lower effort is certain to lose.

Now, assume a contest between two heterogeneous contestants 1 and 2, whose values (or the ability types) are $V_1 > V_2$, implying that contestant 1 is a stronger (or a higher-ranked) contestant. In the lottery model, contestants' efforts (x_i) are given by $x_1 = \frac{V_1^2 V_2}{(V_1+V_2)^2}$, and $x_2 = \frac{V_2^2 V_1}{(V_1+V_2)^2}$, and their probabilities of winning (p_i) are given by $p_1 = \frac{V_1}{V_1+V_2}$ and $p_2 = \frac{V_2}{V_1+V_2}$. In the all-pay case, contestants' efforts are given by $x_1 = \frac{V_2}{2}$, and $x_2 = \frac{{V_2}^2}{2V_1}$, and their probabilities of winning are given by $p_1 = 1 - \frac{V_2}{2V_1}$ and $p_2 = \frac{V_2}{2V_1}$.[16] We can see that these probabilities are derived from contestants' ability types.[17] Therefore, the favorite-underdog type of

[16] For the lottery case, see, for example, Megidish and Sela (2014), who studied two-stage contests that are frequently used in sports competitions. For the all-pay case, see, for example, Krumer and Lechner (2017), who showed that in six out of seven possible cases in Olympic wrestling competitions, the all-pay model predicted correctly the identity of a wrestler with a higher probability of winning.

[17] Note that those probabilities can be easily adjusted for the home advantage. See, for example, Krumer (2013) who provided a theoretical explanation to empirical finding of Page and Page (2007) on second-leg home advantage in the UEFA European Cups, by using the all-pay model adjusted by home and away games.

heterogeneity is the one that fits the economic theory when investigating probabilities of winning (or the number of the gained points per game, in the case of soccer).

3.3 Variables and descriptive statistics

To estimate the effect of playing on non-frequent days on attendance and the number of gained points by the teams, we coded a dummy variable that equals 1 if a match was played on a non-frequent day in a certain league, and zero otherwise. We also used a rich set of variables that characterize team value and players' ability, game attendance, and the international and national schedule. In the following, we present some of the most important measures (a more comprehensive list of variables appears in Appendix A).

Our approach is closely related to that of Krumer and Lechner (2018). Following their study, we used data on players' values from a popular soccer website, *Transfermarkt,* which are supposed to reflect teams' abilities. Since these values increase every season, we standardized them for each league and season so that they take the within-season variation into account.[18] The teams' values measure strongly correlates with teams' performance, suggesting that we have measured teams' abilities quite well.[19] For each game, the favorite is defined as the team with the higher standardized *Transfermarkt* value and the underdog is the team with the lower standardized *Transfermarkt* value. Unlike with betting odds, where favorite and underdog can be a function of the day of the week and the home advantage, the *Transfermarkt* values are determined without considering those factors. Therefore, these definitions are exogenous.

[18] According to Bryson, Frick and Simmons (2013), the coverage of *Transfermarkt* is quite "impressive with information on 190,000 players across 330 football competitions" (p. 611). Players' values are estimated by industry experts and take into account salaries, signing fees, bonuses, and transfer fees. Franck and Nüesch (2012) found that the correlation between values evaluated by *Transfermarkt* and *Kicker*, another highly-respected sport magazine in Germany, is as high as 0.89.

[19] The results of the relevant regression analysis are available upon request from the authors.

Following Krumer (2019), we divided the data into games that take place at the favorite's and the underdog's home fields. In Table 2, which presents descriptive statistics for the pooled data, we can see that when a favorite plays at its home stadium, the average number of points it gains on frequent days is 1.93. When the game is on a non-frequent day, the favorite team gains a very similar number of points (1.98). However, when an underdog team hosts the game, it gains an average of 1.39 points on frequent days and 1.14 points on non-frequent days, suggesting a lower home advantage on non-frequent days for the underdog teams only.

[Insert Table 2 here]

Table 3 presents the descriptive statistics divided into four different leagues, where we can see a similar pattern. However, we also observe that when an underdog team plays at home, in all leagues except La Liga, there are much stronger favorites on non-frequent days compared to frequent days. The direction is the same in La Liga, but the difference between standardized values of favorites on frequent and non-frequent days is less pronounced. This descriptive evidence indicates that there is non-random selection into treatment; that is, non-frequent days. We will discuss how to solve this issue in the next section.

[Insert Table 3 here]

The players' values are used to create additional measures such as the distribution of values between and within teams. More specifically, for each team we compute the standard deviation of players' values − the Herfindahl-Hirschman Index (HHI) − which is defined as the sum of the squares of the values shares of each player within the team. We also created other within-team inequality-related variables such as the ratio of different players' values according to their ranking order in the team. For example, one measure is the ratio between the top three players to players ranked 9−11 according to their values within a team.[20] In addition to players'

[20] See Coates, Frick, and Jewell (2016) for discussion on the relationship between players' inequality in salaries and teams' performance.

values, we also use several other variables that may reflect the level of ability, such as a dummy variable for a team's first season in the top division after being promoted from the lower division, whether a team dismissed its coach during a season, and the age of the coach.[21] We also use data on the size of the squad, share of foreign players in the squad, height of the players, share of left-footed players, age of oldest/youngest players, etc.

Based on the large body of the literature on the effect that the crowd has on home advantage, we created a measure to reflect the attendance in a match. Our first measure − attendance as share of the capacity of the stadium − is the ratio between the number of viewers in a match and the maximal possible capacity of the respective stadium. We also applied a different measure of attendance, namely a natural logarithm of attendance (*Ln(Attendance)*) that is also used in the literature on attendance demand (Buraimo and Simmons, 2015; Buraimo, Tena and de la Piedra, 2018; Krumer, 2019). In addition, following the studies of Boyko, Boyko, and Boyko (2007) and Page and Page (2010), who described the existence of individual differences among referees in terms of the home advantage in different soccer leagues, we created dummy variables for each individual referee in our data.[22] Further, there is also information about the distance between cities, in kilometers for the shortest traveling distance.

We also obtained information on other schedule-related variables in international competitions, such as two pre- and post-World Cup and European Championships months, as well as the months in which the African Cups of Nations and Asian Cup took place. We also took different months of the season and public holidays into account.

[21] See Tena and Forrest (2007), and Flores, Forrest, and Tena (2012) for discussions on the effects of coach dismissals on team performance.

[22] There is also a lot of speculation that some referees have a reputation for being biased in favor or against some teams. See, for example, https://www.sportskeeda.com/slideshow/football-most-biased-referees-top-teams-premier-league. Last accessed on 02.11.2019

4 Empirical Strategy

4.1 Selection into treatment

We studied the effect of playing on a non-frequent day compared to a frequent day on the performance of a team. Here, the challenge for identifying a causal effect lies in the non-random determination of the teams that play at home on non-frequent days. In order to obtain an unbiased causal effect, it is essential to disentangle the effect coming with the selection from the effect caused by playing on non-frequent days. In other words, there is the need to take selection effects into account. Decisions regarding which teams play on which days are made by the calendar committees of the respective leagues and might be driven by teams' characteristics and other schedule-related features, such as public holidays, international breaks, TV broadcasters' interests, European association tournaments, etc..[23]

The rich database presented in the previous section enabled us to opt for a selection-on-observables approach; that is, controlling for the reasons for the deviations from random treatment assignment. Having information on teams and game characteristics, European cups scheduling, national teams' tournaments, etc., enabled us to capture all confounding factors related to team, location, and timing to create a quasi-experimental setup. This allows us to identify the causal effect of playing on non-frequent days on performance if there are no unobserved characteristics that simultaneously affect both the probability of playing on a non-frequent day and the outcome.

An important issue that is worth a separate discussion relates to broadcasting issues, which may affect the allocation of games into frequent and non-frequent days. For instance, a broadcaster may influence the decision to allocate strategically more or less attractive games into different days. However, it is problematic to have a dummy variable of whether a game

[23] See Goossens and Spieksma (2009) for examples of different requirements the calendar committee must consider.

was broadcast or not since it is potentially part of our treatment. We try to solve this issue by capturing the selection effects, which might be partially due to broadcasters' interests, by controlling for a huge range of variables. Among those are the approximated strength of a team and its popularity (in the form of teams' values and teams' dummies, among many other things), as well as controls for other obligations of the teams, so as whether they play in the European tournaments and are therefore shifted to non-frequent days. Therefore, if a broadcaster has a preference for certain teams to play on frequent or non-frequent days (or even if a team itself has such a preference), these teams' dummy variable will capture this potential issue. Thus, it is very likely that selection into treatment is captured by our very large set of potential control variables satisfying the conditional independence assumption needed for credibly estimating a causal effect in our study.

Finally, it is also important to note that the share of capacity (or *Ln(Attendance)*) is an endogenous variable since it is an outcome variable that depends on the day of the game. Therefore, we do not include it as a covariate in our estimation, but only use it as an outcome variable.

4.2 Estimation

4.2.1 Estimator

In order to have a flexible approach and overcome the restrictive assumptions of classical statistical linear models, we used a statistical matching approach. More specifically, we applied the radius-matching-on-the-propensity score estimator with bias adjustment (Lechner, Miquel and Wunsch, 2011).[24] Not only was this estimator found to be very competitive among a range of matching-type estimators, but Huber, Lechner and Wunsch (2013) also showed its superior finite sample and robustness properties in a large-scale Empirical Monte Carlo Study. This

[24] The variance is estimated as weight-based variance as described in Huber, Lechner and Steinmayr (2015), which Bodory et al. (2018) showed to lead to conservative standard errors.

estimator combines the features of distance-weighted radius matching with a bias adjustment, which removes potential biases due to mismatches.[25] Control observations, which are close to the treated unit in terms of the confounding influences, can be compared to the latter to obtain the treatment effect as if treated and control units were in an experimental setting. Therefore, it is crucial to capture all confounding influences; we explain how we do this in more detail below.

4.2.2 Propensity score

The propensity score, which is the probability of playing on a non-frequent day, condenses the information from all relevant confounding variables to a one-dimensional score, determining which observations are similar in terms of confounding influences. In their pioneering work, Rosenbaum and Rubin (1983) showed that controlling for the propensity score removes selection bias. Therefore, treated and non-treated observations with similar propensity scores are compared to each other in the matching estimator.

If the exact relation of confounding variables and the treatment assignment is known, the variables to include in the propensity score estimation can be specified ad-hoc. In our case, we have a set of 385 potentially confounding variables, in addition to the referees and teams dummies, as described in the previous section. Despite prior knowledge about the selection process, we cannot specify ad-hoc exactly which of the many potential confounders to use in the propensity score estimation. Further, including everything would lead to an unfeasible estimation, less precise or instable estimates. Therefore, for the specification we rely on a machine learning algorithm.

Using machine learning for causal inference is not a trivial exercise and the literature is still under development. Since those algorithms are designed for prediction and not for doing inference in treatment effects estimation, we follow the approach of Belloni, Chernozhukov,

[25] Distance weighting leads to a weighting of non-treated observations within the radius inversely proportional to their distance to the respective treated unit.

and Hansen (2014) to make machine learning algorithms useful in this setup. Those authors suggested using the LASSO procedure developed by Tibshirani (1996) as a variable selection tool, twice.[26] [27] In the first step, we selected a set of variables confounding the treatment. In the second step, we selected those variables correlated with the respective outcome.[28] The reason for this double-selection procedure, as opposed to only looking at the treatment selection equation, is to additionally capture variables that are highly correlated to the outcome and mildly related to the treatment selection. The same line of argumentation holds for only looking at the outcome equation. Ignoring those kinds of variables would lead to potentially biased results, as described in Belloni et al. (2014). The union of variables selected by the two separate LASSO procedures is our final set of variables for the propensity score estimation. We repeat this selection procedure for all of the estimations presented in the next section.

5 Results

First, our aim was to study whether playing on non-frequent days has an effect on performance when using the data on all the leagues together. To accomplish this goal, we first estimated the propensity score, based on variables that were chosen in the double-selection LASSO procedure described in the previous section. It is important to note that the purpose of the propensity score estimation is purely technical, to allow the easy purging of the results from the selection effects. Therefore, the respective marginal effects of the propensity score estimation cannot be interpreted in a causal sense, but rather in their contribution to the probability of being treated. Appendix B provides an example of one of the propensity score

[26] The LASSO procedure is a shrinking estimator, which works like an OLS estimator with penalized coefficients. Penalizing the coefficients leads to variables selection as the coefficients of not too informative covariates are forced to zero.

[27] Goller et al. (2019) compared different (machine learning and "classical" probit) estimation procedures for the propensity score in matching estimation and found the LASSO delivered the most credible results in a setup, which is comparable to ours, with many potentially confounding variables and a low share of treated units.

[28] Using the LASSO method requires a penalty term, which is data-driven determined using 10-fold cross-validation. For the current analysis, we chose the penalty term that minimized the mean squared error.

estimations we had to execute, which is the propensity score estimation for the number of favorite's points in games that took place on the favorite's home field. Generally, as is already apparent from Tables 2 and 3, selection effects are driven by team values as well as by schedule-related features such as public holidays and international tournaments. In addition, several individual referees and teams were picked by the double-selection LASSO procedure. While Appendix B only presents the results of the propensity score estimation for the number of points of a favorite team, a separate propensity score estimation for each matching estimation presented in the paper is available upon request.

Table 4 shows the effect of playing on non-frequent days compared to frequent days by pooling all the leagues together. We can see that when a favorite team plays at home, the effect of playing on non-frequent days on the number of points is very close to zero (-0.029) and highly insignificant (p-val=0.64). However, when testing the effect of playing on non-frequent days on the number of points of the favorite when an underdog plays at home, we find that a favorite gains 0.17 points significantly more on non-frequent days than on the frequent days. An underdog gains about 0.13 points less when hosting a game on non-frequent compared to frequent days, making the difference between favorite and underdog about 0.30 points. The share of capacity on non-frequent days is 3.8 percentage points less when a favorite team plays at home and 1.8 percentage points less when an underdog team hosts the game. We observe a similar pattern with *Ln(Attendance)*.

[Insert Table 4 here]

One possible concern is that our results are driven by pooling the data. Therefore, in Table 5, we present the results for each league separately. We can see that, for all the leagues and for both cases of home advantage, we find a lower attendance as a share of capacity and a lower *Ln(Attendance)* on non-frequent days compared to frequent days. When a favorite team hosts the game, the effect of playing on non-frequent days on share of capacity ranges from 1.6 percentage points less in English Premier League to 7.4 percentage points less in German

Bundesliga 1. When looking at *Ln(Attendance)*, we see that a reduction in attendance ranges from 6.5 percent lower attendance in English Premier League to 15.9 percent lower attendance in German Bundesliga 1 on non-frequent days. When an underdog team hosts the game, the effect of playing on non-frequent days on the share of capacity ranges from 3.4 percentage points less in the English Premier League to 5.9 percentage points less in Spanish La Liga. When looking at *Ln(Attendance)*, we see that reduction in attendance ranges from 6.0 percent lower attendance in English Premier League and French Ligue 1 to 10.2 percent lower attendance in German Bundesliga 1. This result replicates the finding of Krumer and Lechner (2018), who also found a lower attendance in the German Bundesliga 1. In addition, although we used a different specification of the days of interest, our results are in line with the findings of Buraimo and Simmons (2015), Buraimo (2008), and Forrest and Simmons (2006), all of whom reported that weekend games attract larger crowds and larger TV ratings in the English Premier League.

[Insert Table 5]

In all the leagues apart from Bundesliga 1, playing on non-frequent days has no effect on the home advantage of the favorite teams. However, in all the four leagues there is a reduced home advantage on non-frequent days for the underdog teams. Our results suggest that the difference in the number of points between the favorite and the underdog teams, when the game takes place on non-frequent days compared to frequent days, is 0.49 in Ligue 1, 0.43 in La Liga, 0.42 in Premier League, and 0.63 in Bundesliga 1. This difference is quite large given that, in our dataset, a favorite with home advantage gains on average about 1.1 points more than the underdog. In addition, in a tight league, one point could make the difference between relegation and survival or between qualification to the UEFA Champions League and the less prestigious UEFA Europa League.

Furthermore, we conducted a WALD test for the equality of each estimate obtained in the analyses for the single leagues. More specifically, we tested the hypothesis that the effects are

equal for all the leagues and reported the p-values from the resulting chi square test statistic in the last column of Table 5. We found that the effects are not statistically different between the leagues for the favorites and underdogs points, or for the *Ln(Attendance).* Although the only significant difference between the leagues is for the share of capacity in the case when a favorite team plays at home, this emphasizes the need to conduct the analysis for each league separately.

Our finding regarding the lower home advantage of the underdog on non-frequent days is in line with the literature on the effect of the density of the crowd and its noise on referees' bias in favor of the home team (Downward and Jones, 2007; Nevill et al., 2002; Page and Page, 2010; Pettersson-Lidbom and Priks, 2010). Therefore, a possible mediator of the difference in the home advantage of the underdog teams in games that take place on non-frequent days is lower crowd noise compared to games on frequent days. Crowd noise might be lower on non-frequent days not only due to lower attendance, but also magnified by a lower amount of alcohol consumption on non-frequent days. In addition, if a crowd is less passionate and supportive on non-frequent days, this may also negatively affect players' motivation, which may result in lower performance.

One possible explanation for the finding that the home advantage of the favorite teams is not affected by the day of the game is that these teams are likely to win because of their higher abilities, regardless of home support. This would suggest that underdog teams depend more on crowd support than favorite teams. The negative effect on the home advantage of the favorite in Bundesliga 1 may be explained by the fact that this league differs from the other leagues as it has 18 teams, not 20. In addition, according to Yi, Goossens, and Nobibon (2019), Bundesliga 1 has the most reliable league schedule, with the lowest number of rescheduled games and a higher quality realized schedule than the other leagues. Indeed, as Table 1 shows, Bundesliga 1 has the lowest amount of games on non-frequent days. In addition, as already mentioned Bundesliga 1 has the highest reduction of attendance among all the leagues when a favorite team plays at home (15.9 percent less attendance). This suggests that fans are not used to

attending the stadium on non-frequent days and also that the players of the favorite teams may be less motivated to play in front of smaller crowds that are also likely to be less supportive on non-frequent days.

6 Conclusion

According to Wright (2014), the main objective of his survey on operational research in sports was fairness, which is probably one of the most important features in sports competitions. In the context of scheduling of the soccer leagues, a schedule would be considered fair if ex-ante all teams have the same probability to convert the home advantage into success, given their individual characteristics, regardless of the day of the game. In this regard, our findings suggest that in all four leagues that we investigated (Bundesliga 1, La Liga, Premier League and Ligue 1), we find that the current schedule structure favors underdog teams that play fewer home games on non-frequent days and favorite teams that play more away games on these days.

Our results also suggest that all four leagues suffer from lower attendance rates on non-frequent days. Therefore, our findings may be of interest to the calendar committees of the relevant leagues, whose task is to allocate games in a way that eliminates any advantage driven by schedule. More specifically, it is important that home games on non-frequent days are allocated evenly on the level of a single team per season. Having said that, we are aware that the calendar committees have to deal with a large amount of constraints, such that it is sometimes almost impossible to satisfy all of them. However, the case of 1.FSV Mainz 05 that was mentioned in the introduction provides an example where the calendar committee disregarded the issue of fair allocation between different days. Furthermore, our findings might be worth an additional restriction, not only in constructing the initial schedule, but also for implementing proactive or reactive strategies for rescheduled games, as Yi, Gosseens, and Nobibon (2019) recently proposed.

In addition, the results of this paper may also help coaches and players prepare to play on different days. According to our results, underdog teams may be expected to have a lower home advantage on non-frequent days and should therefore consider adjusting their preparation to these games. Furthermore, teams may adjust their ticket sales strategy. For example, tickets for games on non-frequent days, for which there is less demand, could be sold for a lower price to attract larger crowds and increase home advantage.

Finally, we call for additional empirical research on different schedule effects in sports leagues that may potentially affect on-pitch performance as well as financial outcomes.

7 References

Belloni, A., Chernozhukov, V. and Hansen, C., 2014. Inference on Treatment Effects After Selection Among High-Dimensional Controls, 81(2) *The Review of Economic Studies* pp. 608−650.

Bodory, H., Camponovo, L., Huber, M. and Lechner, M., 2018. The finite sample performance of inference methods for propensity score matching and weighting estimators. *Journal of Business & Economic Statistics*, forthcoming.

Boyko, R.H., Boyko, A.R. and Boyko, M.G., 2007. Referee bias contributes to home advantage in English Premiership football. Journal of sports sciences, 25(11), pp.1185–1194.

Bryson, A., Frick, B. and Simmons, R., 2013. The returns to scarce talent footedness and player remuneration in European soccer. *Journal of Sports Economics*, 14(6), pp. 606−628.

Buraimo, B., 2008. Stadium attendance and television audience demand in English league football. *Managerial and Decision Economics*, 29(6), pp. 513−523.

Buraimo, B., Migali, G. and Simmons, R., 2016. An analysis of consumer response to corruption: Italy's Calciopoli scandal. *Oxford Bulletin of Economics and Statistics*, 78(1), pp. 22−41.

Buraimo, B. and Simmons, R., 2015. Uncertainty of outcome or star quality? Television audience demand for English Premier League football. *International Journal of the Economics of Business*, 22(3), pp. 449−469.

Buraimo, B., Tena, J.D. and de la Piedra, J.D., 2018. Attendance demand in a developing football market: the case of the Peruvian first division. *European Sport Management Quarterly*, 18(5), pp. 671–686.

Coates, D. Frick, B. and Jewell, T., 2016. Superstar salaries and soccer success: The impact of designated players in Major League Soccer. *Journal of Sports Economics*, 17(7), pp. 716−735.

Della Croce, F. and Oliveri, D., 2006. Scheduling the Italian football league: An ILP-based approach. *Computers & Operations Research*, 33(7), pp. 1963−1974.

Downward, P. and Jones, M., 2007. Effects of crowd size on referee decisions: Analysis of the FA Cup. *Journal of Sports Sciences*, 25(14), pp. 1541−1545.

Durán, G., Durán, S., Marenco, J., Mascialino, F. and Rey, P.A., 2019. Scheduling Argentina's professional basketball leagues: A variation on the Travelling Tournament Problem. *European Journal of Operational Research*, 275(3), pp. 1126–1138.

Durán, G., Guajardo, M. and Sauré, D., 2017. Scheduling the South American Qualifiers to the 2018 FIFA World Cup by integer programming. *European Journal of Operational Research*, 262(3), pp. 1109–1115.

Flores, R., Forrest, D. and Tena, J.D., 2012. Decision taking under pressure: Evidence on football manager dismissals in Argentina and their consequences. *European Journal of Operational Research*, 222(3), pp. 653−662.

Forrest, D. and Simmons, R., 2006. New issues in attendance demand the case of the English Football League. *Journal of Sports Economics*, 7(3), pp. 247−266.

Franck, E. and Nüesch, S., 2012. Talent and/or popularity: What does it take to be a superstar? *Economic Inquiry*, 50(1), pp. 202−216.

Goller, D., Lechner, M., Moczall, A. and Wolff, J., 2019. Does the Estimation of the Propensity Score by Machine Learning improve Matching Estimation? The case of Germany's Programmes for Long Term Unemployed. *(No. 12526) Institute of Labor Economics (IZA).*

Goossens, D. and Spieksma, F., 2009. Scheduling the Belgian soccer league. *Interfaces*, 39(2), pp. 109−118.

Goossens, D. and Spieksma, F., 2012. Soccer schedules in Europe: an overview. *Journal of Scheduling*, 15(5), pp. 641−651.

Huber, M., Lechner, M. and Steinmayr, A., 2015. Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour. *Empirical Economics*, 49(1), pp. 1−31.

Huber, M., Lechner, M. and Wunsch, C., 2013. The performance of estimators based on the propensity score. *Journal of Econometrics*, 175(1), pp. 1−21.

Kendall, G., 2008. Scheduling English football fixtures over holiday periods. *Journal of the Operational Research Society*, 59(6), pp. 743−755.

Kendall, G., Knust, S., Ribeiro, C.C. and Urrutia, S., 2010. Scheduling in sports: An annotated bibliography. *Computers & Operations Research*, 37(1), pp. 1−19.

Kendall, G. and Lenten, L.J., 2017. When sports rules go awry. *European Journal of Operational Research*, 257(2), pp. 377−394.

Klumpp, T. and Polborn, M.K., 2006. Primaries and the New Hampshire effect. *Journal of Public Economics*, 90(6−7), pp. 1073−1114.

Krumer, A., 2013. Best-of-two contests with psychological effects. *Theory and Decision*, 75(1), pp. 85−100.

Krumer, A., 2019. Testing the effect of kick-off time in the UEFA Europa League. *European Sport Management Quarterly*, forthcoming.

Krumer, A., Megidish, R. and Sela, A., 2019. The optimal design of round-robin tournaments with three players. *Journal of Scheduling*, forthcoming.

Krumer, A. and Lechner, M., 2017. First in first win: Evidence on schedule effects in round-robin tournaments in mega-events. *European Economic Review*, 100, pp. 412−427.

Krumer, A. and Lechner, M., 2018. Midweek effect on soccer performance: Evidence from the German Bundesliga. *Economic Inquiry*, 56(1), pp. 193−207.

Lechner, M., Miquel, R. and Wunsch, C., 2011. Long-run effects of public sector sponsored training in West Germany. *Journal of the European Economic Association*, 9(4), pp. 721−784.

Megidish, R. and Sela, A., 2014. Sequential contests with synergy and budget constraints. *Social Choice and Welfare*, 42(1), pp. 215−243.

Nevill, A.M., Balmer, N.J. and Williams, A.M., 2002. The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4), pp. 261−272.

Page, L. and Page, K., 2007. The second leg home advantage: Evidence from European football cup competitions. *Journal of Sports Sciences*, 25(14), pp. 1547−1556.

Page, K. and Page, L., 2010. Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology*, 31(2), pp. 192−199.

Pettersson-Lidbom, P. and Priks, M., 2010. Behavior under social pressure: Empty Italian stadiums and referee bias. *Economics Letters*, 108(2), pp. 212−214.

Rasmussen, R.V. and Trick, M.A., 2008. Round robin scheduling – a survey. *European Journal of Operational Research*, 188(3), pp. 617−636.

Ribeiro, C.C. and Urrutia, S., 2007. Heuristics for the mirrored traveling tournament problem. *European Journal of Operational Research*, 179(3), pp. 775−787.

Rohde, M. and Breuer, C., 2017. Financial Incentives and Strategic Behavior in European Professional Football: A Match Day Analysis of Starting Squads in the German Bundesliga and UEFA Competitions. *International Journal of Sport Finance*, 12(2).

Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), pp. 41−55.

Ponzo, M. and Scoppa, V., 2018. Does the home advantage depend on crowd support? Evidence from same-stadium derbies. *Journal of Sports Economics*, 19(4), pp. 562–582.

Szymanski, S., 2003. The economic design of sporting contests. *Journal of Economic Literature*, 41(4), pp. 1137−1187.

Tena, J.D. and Forrest, D., 2007. Within-season dismissal of football coaches: Statistical analysis of causes and consequences. *European Journal of Operational Research*, 181(1), pp. 362−373.

Tibshirani, R., 1996. Regression Shrinkage and Selection via the lasso". *Journal of the Royal Statistical Society*, Series B, 58(1), pp. 267–88.

Tullock, G., 1980. Efficient rent-seeking. In J.M. Buchanan, R.D. Tollison and G. Tullock (Eds.), Toward a theory of the rent-seeking society (pp. 97−112). College Station: Texas A&M University Press

Van Bulck, D., Goossens, D., Schönberger, J. and Guajardo, M., 2020. RobinX: A three-field classification and unified data format for round-robin sports timetabling. *European Journal of Operational Research*, 280(2), pp. 568–580.

Wang, C., Goossens, D. and Vandebroek, M., 2018. The impact of the soccer schedule on TV viewership and stadium attendance: evidence from the Belgian Pro League. *Journal of Sports Economics*, 19(1), pp. 82−112.

Wright, M., 2014. OR analysis of sporting rules – A survey. *European Journal of Operational Research*, 232(1), pp. 1−8.

Yi, X., Goossens, D. and Nobibon, F.T., 2019. Proactive and Reactive Strategies for Football League Timetabling. *European Journal of Operational Research*, forthcoming.

Table 1: Frequency of matches by days

Weekday	Ligue 1	La Liga	Premier League	Bundesliga
Monday	**11**	199	**204**	**1**
Tuesday	**57**	**71**	**182**	**72**
Wednesday	251	**185**	287	**96**
Thursday	**14**	**70**	**26**	**3**
Friday	**186**	**140**	**30**	275
Saturday	2188	1285	2152	1994
Sunday	1093	1850	919	619

Notes: The total numbers of matches on the specific weekdays, including matches that were excluded from the final analysis. Bold numbers represent the days of the "non-frequent" specification for each league.

Table 2: Descriptive statistics of the main variables

	Favorite playing home		Underdog playing home	
Variable	**Frequent days**	**Non-frequent days**	**Frequent days**	**Non-frequent days**
Game Outcomes				
Favorite points	1.930 (1.249)	1.981 (1.245)	1.333 (1.292)	1.575 (1.298)
Underdog points	0.817 (1.131)	0.784 (1.126)	1.391 (1.298)	1.144 (1.251)
Favorite win	0.559	0.582	0.352	0.432
Underdog win	0.188	0.183	0.371	0.288
Game Characteristics				
Visitors	32,743 (17,533)	34,583 (18,871)	24,031 (12,640)	24,943 (12,584)
Stadium capacity	40,825 (17,510)	44,864 (19,521)	29,924 (12,974)	31,410 (13,588)
Share of capacity	0.781 (0.192)	0.750 (0.201)	0.792 (0.183)	0.794 (0.176)
Ln(Attendance)	10.242 (0.584)	10.292 (0.604)	9.955 (0.529)	10.005 (0.497)
Distance (in km)	487.5 (341.8)	501.6 (364.6)	485.1 (337.8)	461.5 (402.3)
Teams Characteristics				
Fav. standardized team value	0.218 (1.009)	0.464 (1.143)	0.228 (1.007)	0.510 (1.198)
Und. standardized team value	-0.577 (0.352)	-0.508 (0.333)	-0.570 (0.357)	-0.515 (0.391)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.410 (0.713)	2.492 (0.826)	2.408 (0.727)	2.469 (0.737)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.330 (0.750)	2.403 (0.815)	2.332 (0.753)	2.381 (0.754)
Fav. plays in Europa League or Champions League	0.331	0.455	0.340	0.440
Und. plays in Europa League or Champions League	0.086	0.127	0.084	0.108
Schedule-related				
African Cup of Nations	0.066	0.023	0.066	0.037
Asian Cup	0.022	0.009	0.021	0.007
Public holiday	0.035	0.091	0.035	0.091
2 months before UEFA European Championship	0.064	0.084	0.063	0.073
2 months after UEFA European Championship	0.042	0.065	0.041	0.048
2 months before FIFA World Cup	0.040	0.039	0.044	0.027
2 months after FIFA World Cup	0.028	0.049	0.028	0.044
Observations	4462	569	4548	563

Notes: This table presents average values and standard deviations (in parentheses for non-binary variables) for the main variables from all league combined. Descriptive statistics for all available variables appear in Appendix A.

Table 3: Descriptive statistics of the main variables for each league separately

Variable	Favorite playing home		Underdog playing home	
	Frequent days	**Non-frequent days**	**Frequent days**	**Non-frequent days**
		Ligue 1		
Favorite points	1.905 (1.229)	1.901 (1.221)	1.309 (1.286)	1.685 (1.266)
Underdog points	0.807 (1.093)	0.793 (1.071)	1.410 (1.297)	1.000 (1.177)
Share of capacity	0.649 (0.183)	0.634 (0.210)	0.691 (0.178)	0.728 (0.181)
Ln(Attendance)	9.859 (0.566)	10.052 (0.592)	9.604 (0.444)	9.821 (0.408)
Distance (in km)	657.2 (335.3)	662.1 (308.0)	658.8 (331.0)	608.2 (339.2)
Fav. standardized team value	0.310 (1.102)	0.769 (1.313)	0.303 (1.087)	1.127 (1.519)
Und. standardized team value	-0.542 (0.357)	-0.405 (0.286)	-0.535 (0.367)	-0.426 (0.241)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.388 (0.779)	2.754 (1.183)	2.406 (0.823)	2.508 (0.855)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.303 (0.756)	2.329 (0.695)	2.297 (0.751)	2.277 (0.556)
Fav. plays in Europa League or Champions League	0.273	0.378	0.293	0.293
Und. plays in Europa League or Champions League	0.079	0.108	0.074	0.054
African Cup of Nations	0.060	0.036	0.050	0.043
Public holiday	0.032	0.054	0.026	0.022
Weekdays	Wed, Sat, Sun	Mon, Tue, Thu, Fri	Wed, Sat, Sun	Mon, Tue, Thu, Fri
Observations	1365	111	1397	92
		La Liga		
Favorite points	2.005 (1.249)	2.071 (1.227)	1.307 (1.300)	1.446 (1.281)
Underdog points	0.777 (1.137)	0.717 (1.103)	1.432 (1.313)	1.246 (1.256)
Share of capacity	0.698 (0.169)	0.651 (0.160)	0.703 (0.178)	0.680 (0.165)
Ln(Attendance)	10.154 (0.610)	10.077 (0.636)	9.763 (0.477)	9.722 (0.481)
Distance (in km)	631.2 (401.2)	659.0 (407.2)	621.5 (389.1)	688.2 (503.8)
Fav. standardized team value	0.193 (1.010)	0.234 (1.073)	0.190 (1.012)	0.224 (1.077)
Und. standardized team value	-0.520 (0.257)	-0.499 (0.237)	-0.518 (0.246)	-0.510 (0.231)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.614 (0.779)	2.630 (0.809)	2.604 (0.779)	2.624 (0.840)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.450 (0.822)	2.525 (0.903)	2.457 (0.827)	2.522 (0.886)
Fav. plays in Europa League or Champions League	0.328	0.381	0.323	0.390
Und. plays in Europa League or Champions League	0.087	0.115	0.083	0.092
African Cup of Nations	0.074	0.000	0.087	0.015
Public holiday	0.043	0.013	0.045	0.015
Weekdays	Mon, Sat, Sun	Tue, Wed, Thu, Fri	Mon, Sat, Sun	Tue, Wed, Thu, Fri
Observations	1078	226	1105	195

Table 3, continued

Variable	Favorite playing home		Underdog playing home	
	frequent days	**non-frequent days**	**frequent days**	**non-frequent days**
		Premier League		
Favorite points	1.968 (1.244)	2.033 (1.216)	1.389 (1.289)	1.600 (1.292)
Underdog points	0.790 (1.124)	0.719 (1.073)	1.322 (1.281)	1.111 (1.236)
Share of capacity	0.927 (0.085)	0.936 (0.083)	0.913 (0.091)	0.906 (0.095)
Ln(Attendance)	10.486 (0.364)	10.599 (0.372)	10.199 (0.318)	10.205 (0.357)
Distance (in km)	235.8 (132.7)	203.3 (131.2)	233.6 (133.2)	208.0 (126.8)
Fav. standardized team value	0.213 (0.990)	0.542 (0.998)	0.246 (0.984)	0.552 (1.099)
Und. standardized team value	-0.654 (0.409)	-0.596 (0.412)	-0.636 (0.419)	-0.562 (0.549)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.200 (0.514)	2.223 (0.473)	2.201 (0.512)	2.210 (0.484)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.336 (0.737)	2.336 (0.813)	2.326 (0.734)	2.383 (0.756)
Fav. plays in Europa League or Champions League	0.385	0.516	0.398	0.511
Und. plays in Europa League or Champions League	0.093	0.137	0.099	0.153
African Cup of Nations	0.060	0.059	0.060	0.074
Public holiday	0.039	0.281	0.044	0.242
Weekdays	Wed, Sat, Sun	Mon, Tue, Thu, Fri	Wed, Sat, Sun	Mon, Tue, Thu, Fri
Observations	1111	153	1124	190
		Bundesliga 1		
Favorite points	1.833 (1.279)	1.734 (1.365)	1.331 (1.295)	1.698 (1.381)
Underdog points	0.914 (1.184)	1.089 (1.322)	1.396 (1.302)	1.140 (1.347)
Share of capacity	0.898 (0.115)	0.837 (0.144)	0.902 (0.125)	0.880 (0.133)
Ln(Attendance)	10.624 (0.378)	10.645 (0.448)	10.416 (0.414)	10.401 (0.420)
Distance (in km)	369.8 (183.1)	403.3 (171.7)	365.0 (185.4)	350.4 (185.2)
Fav. standardized team value	0.116 (0.865)	0.543 (1.231)	0.139 (0.891)	0.406 (1.036)
Und. standardized team value	-0.604 (0.345)	-0.512 (0.411)	-0.603 (0.357)	-0.517 (0.377)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.457 (0.661)	2.254 (0.566)	2.428 (0.661)	2.648 (0.662)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.222 (0.641)	2.291 (0.667)	2.243 (0.661)	2.168 (0.511)
Fav. plays in Europa League or Champions League	0.352	0.658	0.360	0.558
Und. plays in Europa League or Champions League	0.087	0.165	0.084	0.105
African Cup of Nations	0.072	0.000	0.073	0.000
Public holiday	0.024	0.000	0.027	0.000
Weekdays	Fri, Sat, Sun	Mon, Tue, Wed, Thu	Fri, Sat, Sun	Mon, Tue, Wed, Thu
Observations	908	79	922	86

Notes: This table presents average values and standard errors (in parentheses for non-binary variables) for the main variables for each league separately. For each league, we also provide the used frequent and non-frequent weekdays.

Table 4: Levels and effects of playing on non-frequent days for all the data

Dependent Variables	Exp. value on non-frequent days (1)	Exp. value on frequent days (2)	Effect of playing on non-frequent days (3)	Standard error of the effect (4)	Common support (in %) (5)
		Favorite team playing home			
Favorites Points	1.911	1.940	-0.029	0.062	98.5%
Underdogs Points	0.826	0.809	0.017	0.057	96.9%
Share of capacity	0.745	0.783	-0.038***	0.010	93.3%
ln(Attendance)	10.185	10.258	-0.073**	0.032	97.5%
		Underdog team playing home			
Favorites Points	1.499	1.329	0.170**	0.063	99.9%
Underdogs Points	1.251	1.376	-0.125***	0.062	99.4%
Share of capacity	0.775	0.793	-0.018**	0.009	99.9%
ln(Attendance)	9.903	9.961	-0.058**	0.024	99.8%

Notes: The results represent all of the data. Columns (1) and (2) represent the expected values for non-frequent and frequent days, respectively. Columns (3) and (4) report the average treatment effect and the respective standard errors. Standard errors are calculated as weight-based standard errors and clustered at the season per league level. Column (5) states the share of observations in common support in the radius matching. ** and *** denote the 5% and 1% significance levels, respectively.

Table 5: Levels and Effects of playing on non-frequent days for each league separately

Outcomes	Ligue 1	La Liga	Premier League	Bundesliga 1	Wald Test p-value
		Favorite team playing home			
Favorites Points	-0.062	0.015	-0.093	-0.276	0.55
	(0.124)	(0.111)	(0.117)	(0.169)	
Underdogs Points	-0.064	0.135	-0.046	0.310*	0.17
	(0.113)	(0.095)	(0.103)	(0.168)	
Share of capacity	-0.042*	-0.066***	-0.016*	-0.074***	0.00
	(0.025)	(0.014)	(0.009)	(0.016)	
Ln(Attendance)	-0.102*	-0.116**	-0.065*	-0.159***	0.50
	(0.056)	(0.052)	(0.034)	(0.053)	
		Underdog team playing home			
Favorites Points	0.228*	0.242**	0.225**	0.329**	0.96
	(0.128)	(0.108)	(0.114)	(0.170)	
Underdogs Points	-0.266**	-0.188*	-0.197*	-0.303*	0.91
	(0.138)	(0.100)	(0.110)	(0.162)	
Share of capacity	-0.036**	-0.059***	-0.034***	-0.036**	0.44
	(0.018)	(0.013)	(0.009)	(0.016)	
Ln(Attendance)	-0.060*	-0.073*	-0.060**	-0.102**	0.90
	(0.035)	(0.041)	(0.029)	(0.051)	

Notes: The results represent the effects of playing on non-frequent days for each league separately. Standard errors, as presented in parentheses, are calculated as weight-based standard errors and clustered at the seasonal level. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively. Common support for each of the matching estimations is at least 87% (Ligue 1), 92% (La Liga), 91% (Premier League), and 84% (Bundesliga).

Appendix A: Descriptive statistics

Table A.1: Descriptive statistics for all available variables

	Favorite playing home		Underdog playing home	
Variable	**frequent days**	**non-frequent days**	**frequent days**	**non-frequent days**
Game Outcomes				
Favorite points	1.930	1.981	1.333	1.575
Underdog points	0.817	0.784	1.391	1.144
Favorite win	0.559	0.582	0.352	0.432
Underdog win	0.188	0.183	0.371	0.288
Game Characteristics				
Stadium capacity	40,825	44,864	29,924	31,410
Visitors	32,743	34,583	24,031	24,943
Share of capacity	0.781	0.750	0.792	0.794
Ln(Attendance)	10.242	10.292	9.955	10.005
Distance	487.502	501.569	485.106	461.455
Teams Characteristics				
Fav. team value (in mill. €)	147.424	194.466	148.308	204.742
Und. team value (in mill. €)	63.947	73.319	64.615	78.651
Fav. standardized team value	0.218	0.464	0.228	0.510
Und. standardized team value	-0.577	-0.508	-0.570	-0.515
Fav. mean value (in mill. €)	4.389	5.635	4.407	5.893
Und. mean value (in mill. €)	1.933	2.209	1.955	2.336
Fav. median value (in mill. €)	2.999	3.738	3.012	3.964
Und. median value (in Mill €)	1.473	1.632	1.479	1.760
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.410	2.492	2.408	2.469
Und. ratio of top 3 to ranked 9−11 most valuable players	2.330	2.403	2.332	2.381
Fav. plays in Champions League or Europa League	0.331	0.455	0.340	0.440
Und. plays in Champions League or Europa League	0.086	0.127	0.084	0.108
Fav. plays in Europa League	0.176	0.221	0.183	0.211
Und. plays in Europa League	0.070	0.112	0.066	0.083
Fav. plays in Champions League	0.184	0.290	0.188	0.281
Und. plays in Champions League	0.023	0.026	0.024	0.041
Newcomer	0.046	0.054	0.292	0.242
Fav. coaches age	49.280	49.708	49.325	49.917
Und. coaches age	48.422	48.815	48.446	49.110
Fav. first match with new coach	0.024	0.018	0.027	0.014
Und. first match with new coach	0.024	0.011	0.021	0.021
Fav. second match with new coach	0.025	0.021	0.023	0.021
Und. second match with new coach	0.023	0.039	0.025	0.023

Table A.1 continued

Variable	Favorite playing home		Underdog playing home	
	frequent days	**non-frequent days**	**frequent days**	**non-frequent days**
Teams Characteristics				
Fav. third match with new coach	0.022	0.019	0.024	0.020
Und. third match with new coach	0.024	0.016	0.025	0.023
Fav. fourth match with new coach	0.021	0.025	0.021	0.036
Und. fourth match with new coach	0.020	0.023	0.021	0.027
Fav. fifth match with new coach	0.023	0.035	0.024	0.032
Und. fifth match with new coach	0.016	0.026	0.019	0.032
Fav. size of squad	33.363	34.186	33.384	34.561
Und. size of squad	32.753	32.889	32.704	33.286
Fav. number of foreigners	17.525	17.427	17.540	18.078
Und. number of foreigners	15.822	15.202	15.886	16.089
Fav. share of foreigners	0.520	0.503	0.520	0.515
Und. share of foreigners	0.478	0.457	0.479	0.476
Fav. share right-footed	0.696	0.692	0.696	0.680
Und. share right-footed	0.694	0.709	0.695	0.695
Fav. share left-footed	0.213	0.206	0.212	0.219
Und. share left-footed	0.222	0.218	0.219	0.225
Fav. share both-footed	0.073	0.084	0.073	0.089
Und. share both-footed	0.061	0.054	0.063	0.063
Fav. mean height (in cm)	181.816	181.453	181.785	181.677
Und. mean height (in cm)	181.778	181.534	181.786	181.534
Fav. min height (in cm)	171.736	171.170	171.715	171.391
Und. min height (in cm)	172.026	171.525	171.982	171.666
Fav. max height (in cm)	191.623	191.817	191.586	191.933
Und. max height (in cm)	191.005	191.111	191.039	191.206
Fav. std.dev. height	6.288	6.562	6.295	6.520
Und. std.dev. height	5.986	6.116	6.003	6.147
Fav. HHI	0.060	0.061	0.060	0.060
Und. HHI	0.058	0.059	0.058	0.058
Fav. std.dev. HHI	0.031	0.032	0.031	0.032
Und. std.dev. HHI	0.028	0.029	0.028	0.029
Fav. ratio of top 3 to ranked 12−14 most valuable player	3.065	3.181	3.065	3.135
Und. ratio of top 3 to ranked 12−14 most valuable player	2.927	3.044	2.926	3.002
Fav. ratio of top 11 to ranked 12−23 most valuable player	2.879	2.939	2.881	2.879
Und. ratio of top 11 to ranked 12−23 most valuable player	2.704	2.779	2.702	2.690

Table A.1 continued

Variable	Favorite playing home		Underdog playing home	
	frequent days	**non-frequent days**	**frequent days**	**non-frequent days**
Teams Characteristics				
Fav. median age	23.896	23.927	23.890	23.915
Und. median age	24.356	24.502	24.360	24.571
Fav. mean age	24.258	24.282	24.250	24.265
Und. mean age	24.663	24.762	24.666	24.813
Fav. std.dev. age	4.470	4.502	4.477	4.533
Und. std.dev. age	4.453	4.430	4.455	4.434
Fav. mean age 11 most valuable player	25.868	25.825	25.878	25.780
Und. mean age 11 most valuable player	26.106	26.024	26.118	26.027
Fav. age ratio of top 11 to ranked 12−23 most valuable players	1.006	1.000	1.006	0.995
Und. age ratio of top 11 to ranked 12−23 most valuable players	1.003	0.994	1.003	0.994
Fav. mean age if aged above 20	26.069	26.111	26.076	26.151
Und. mean age if aged above 20	26.239	26.265	26.240	26.297
Fav. min age	17.239	17.146	17.225	17.091
Und. min age	17.401	17.420	17.413	17.414
Fav. max age	34.232	34.330	34.239	34.433
Und. max age	34.455	34.591	34.501	34.671
Schedule-related				
African Cup of Nations	0.066	0.023	0.066	0.037
Asian Cup	0.022	0.009	0.021	0.007
Public holiday	0.035	0.091	0.035	0.091
Christmas holiday (24.12.−01.01.)	0.015	0.095	0.015	0.105
After international break	0.123	0.035	0.116	0.048
Before UEFA European Championship	0.064	0.084	0.063	0.073
After UEFA European Championship	0.042	0.065	0.041	0.048
Before FIFA World Cup	0.040	0.039	0.044	0.027
After FIFA World Cup	0.028	0.049	0.028	0.044
Observations	4462	569	4548	563

Notes: This table presents average values for all available variables from all league combined. *Und.* and *Fav.* represent the coefficients for the underdog and favorite teams, respectively.

Appendix B: Propensity score estimation

Table B.1: Estimation of propensity score

Variables	Favorite playing home	Underdog playing home
Schedule and match characteristics		
Season 2009/10	-0.029*	
Season 2013/14	0.027*	
Season 2014/15	0.044***	0.035**
Season 2015/16	0.031**	0.055***
Ligue 1	-0.056**	-0.068***
Premier League		0.045**
March	0.011	
August	-0.065***	-0.082***
November		-0.034*
After international break	-0.104***	
Public holiday	0.103***	0.085***
Referee 28	0.131	
Referee 125		-0.042
Referee 178		-0.011
Referee 220		0.105
Team characteristics		
Fav. standardized team value	0.007	0.016
Und. standardized team value	0.046**	0.045*
Underdog team value (in mill. €)	-0.000	-0.000
Favorite average team value (in mill. €)	-0.004	0.003
Fav. std.dev. team value (in mill. €)	0.006	
Und. ratio of value of top 3 to ranked 9−11 most valuable players		0.007
Und. ratio of value of top 11 to ranked 12−21 most valuable players		-0.005
Und. Athletic Bilbao	-0.009	
Und. Getafe CF	0.039	
Und. EA Guingamp	-0.020	
Und. RC Mallorca	0.014	
Fav. OG Nice	0.042	
Und. VfB Stuttgart	0.016	
Und. Deportivo Alaves		-0.001
Fav. Tottenham Hotspurs		-0.044
Und. Xerez CD		-0.048
Und. Eintracht Frankfurt		0.042
Fav. UD Las Palmas		0.157
Fav. Manchester United		-0.045
Und. Manchester United		0.235*
Und. Ajaccio GFCO	-0.033	
Fav. Cordoba CF	0.113	
Fav. plays in Europa League	0.026**	
Und. plays in Europa League	0.020	
Und. plays in Champions League	-0.056*	-0.000
Und. std.dev. height of 11 most valuable players	0.005	
Fav. age of 11 most valuable players	-0.001	-0.003
Und. maximum height of 11 most valuable players	-0.002	
Und. coach age	0.001	0.001**
Fav. squad size		0.006***
Fav. number of foreigners in squad		-0.003**
Und. share of foreigners	-0.086**	-0.061*
Number of observations	5031	5111

Notes: Dependent variable is whether a game is played on non-frequent day. Probit average marginal effects are presented. The results are based on the union of variables selected by the two-step LASSO variable selection for playing on non-frequent days and the number of favorites' points. *Und.* and *Fav.* represent the underdog and favorite teams, respectively. *, ** and *** represent the 10%, 5%, and 1% significance levels, respectively.

Appendix C: List of sources for database

www.uefa.com

www.fifa.com

www.transfermarkt.com

www.football-data.co.uk

www.rsssf.com

www.espnfc.com

www.fcal.ch

https://en.wikipedia.org/wiki/Bundesliga

www.weltfussball.com

www.google.com/maps

www.kicker.de